AD/A-001 534

DEVELOPMENT OF A PROGRAMED TESTING SYSTEM

A. G. Bayroff, et al

Army Research Institute for the Behavioral and Social Sciences
Arlington, Virginia

December 1974

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Paper 259	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD/A-001 534
4. TITLE (and Subtitle) DEVELOPMENT OF A PROGRAMED TESTING SYSTEM		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A. G. Bayroff, Robert M. Ross, and M. A. Fischl		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS U. S. Army Research Institute for the Behavioral and Social Sciences 1300 Wilson Blvd., Arlington, VA. 22209		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2T061101A91B
11. CONTROLLING OFFICE NAME AND ADDRESS Office of the Deputy Chief of Staff for Personnel, Washington, DC 20310		12. REPORT DATE December 1974 13. NUMBER OF PAGES 21
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield VA 22151

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Automated testing
Computer assisted testing
Tailored testing
Branching tests
Machine administered tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes an automated system for administering, scoring, and recording results of multiple-choice tests.

The system consists of examinee station, proctor station, and central computer; the report describes the equipment and the programing characteristics of the respective components. The system is designed for tests tailored to the ability of the examinee, in which a more difficult question is presented after each correct answer or an easier question after each incorrect answer.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

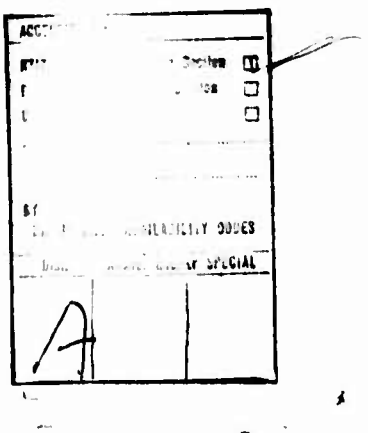
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

J. E. UHLANER
Technical Director

R. A. ROOTH
COL, GS
Commander



NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-P, 1300 Wilson Boulevard, Arlington, Virginia 22209.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Technical Paper 259

AD

DEVELOPMENT OF A PROGRAMED TESTING SYSTEM

A. G. Bayroff, Robert M. Ross, and M. A. Fischl

INDIVIDUAL TRAINING AND MANPOWER DEVELOPMENT TECHNICAL AREA
Edmund F. Fuchs, Chief

Submitted By:
E. Ralph Dusek, Director
INDIVIDUAL TRAINING AND PERFORMANCE
RESEARCH LABORATORY

Approved By:
J. E. Uhlaner
TECHNICAL DIRECTOR

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

Office, Deputy Chief of Staff for Personnel
Department of the Army
1300 Wilson Boulevard, Arlington, Virginia 22209

December 1974

Army Project Number
2T061101A91B

Recruitment Systems &
Manpower Requirements

Approved for public release; distribution unlimited

ARI Research Reports and Technical Papers are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The potential of computer-programed testing systems has only begun to be exploited. Two studies by the Army Research Institute for the Behavioral and Social Sciences (ARI) had indicated the promise of programed testing and test machines. Research on branching tests, in which the item sequence becomes a function of the pattern of correct/incorrect responses elicited from the examinee, indicated that branching tests offer the prospect of increased reliability per unit of testing time in comparison with conventional tests. In addition, research on the feasibility of constructing a machine to present test items, record and score responses, and determine the next item for presentation indicated that such a device is completely within the state of the art.

Accordingly, an interim system was developed, to serve as the pilot for an eventual machine testing system while providing the means for further research on branching tests. This interim system utilizes in-house/off-the-shelf capability, with its basis in the ARI computer and peripheral equipment from the ARI Information Systems Laboratory. The entire project is responsive to requirements of RDTE Project 2T061101A91B, "Computerized Tests for AFEES Screening," FY 1974 Work Program.

J. E. UHLANER
Technical Director

DEVELOPMENT OF A PROGRAMED TESTING SYSTEM

BRIEF

Requirement:

To develop a fully automated prototype testing system for administering, scoring, and recording results of multiple-choice tests.

Research Product:

The interim testing system consists of an examinee station with a projection screen, a pushbutton panel, and CRT; a proctor station with message keyboard and CRT; and the control computer. In use, a question is projected onto the examinee's screen, its multiple choices aligned with a column of labeled pushbuttons. The examinee pushes the button directly opposite his choice of answer; he then pushes a second button labeled RECORD to finalize his answer in the computer. If his answer is correct, the computer presents a more difficult question; if not, he is given an easier one. Testing proceeds at the examinee's pace, within administrative limits. The proctor has only emergency duties once testing begins. The system described is an off-the-shelf model utilizing the ARI computer.

Utilization:

An automated, programed testing system would permit not only a greatly reduced administrative staff but tests which were adapted to an individual examinee. Such individualization offers the prospect of greater reliability per unit of testing time, as a result of matching the test more closely to the ability of the examinee and thus reducing measurement errors due to carelessness on too easy items or lucky guessing on difficult ones. The interim system will serve as a pilot for future programed and machine testing.

DEVELOPMENT OF A PROGRAMED TESTING SYSTEM

CONTENTS

	Page
SYSTEM OVERVIEW	2
Requirements	2
System Description	2
COMPUTER PROGRAMING	6
Response Analysis	9
SUMMARY OF PROGRAMED TESTING SYSTEM CAPABILITIES	10
DISTRIBUTION LIST	12

FIGURES

Figure 1.	Examinee station	4
2.	Schematic representation of the programed testing system	5
3.	Flow chart--operation of a single testing station	7

UA

DEVELOPMENT OF A PROGRAMED TESTING SYSTEM

Two Army Research Institute (ARI) studies conducted in the last ten years have indicated the promise of programed testing and test machines. Branching tests, in which the item sequence becomes a function of the pattern of correct/incorrect responses elicited from the examinee, were studied by Bayroff and Seeley [1]. Their research indicated that branching tests, in comparison with conventional tests, offered the prospect of increased reliability per unit of testing time as a result of the individualization of the test to the ability level of every examinee. That is, measurement errors attributable to such factors as an examinee's clerical mistakes on items much too easy for him, or correct guessing on items much too hard for him, would be greatly reduced because the branching procedure would expose the examinee to a minimum of items so disparate from his ability level.

In addition to the branching research, Bayroff had studied the feasibility of constructing a machine to present test items, record and score responses, and determine the next item for presentation [2]. The study had indicated that construction of such a device was completely within the state of the art.

Accordingly, it was decided to attempt to develop an interim system, which would serve as the pilot for an eventual machine testing system while at the same time providing the means for further research on branching tests. The intention was to develop this interim system utilizing as much in-house/off-the-shelf capability as possible. This meant that the system would have its basis in the ARI computer, and would utilize peripheral equipment from the ARI Information Systems Laboratory. The complete development required selection of equipment components, determination of necessary modifications to them, preparation of test items in their presentation medium, provision of means for input to and output from the computer, and composition of a program to integrate all of the components into a system.

---

[1] Bayroff, A. G. and L. C. Seeley. An exploratory study of branching tests. ARI Technical Research Note 188. June 1967.

[2] Bayroff, A. G. Feasibility of a programed testing machine. ARI Research Study 64-3. November 1964.

SYSTEM OVERVIEW

Requirements

A set of requirements for a programed testing system, identified in a previous report,[3] is summarized here to provide a focus and perspective.

In general, a programed testing system must select a test item from its pool, present the item to an examinee, also present all necessary instructions, accept and score responses, and direct the next item to be presented. To accomplish these things:

(1) The system should work completely automatically, without the necessity for human staffing in any role except as administrative supervisor to greet examinees, assure that they are established at the machine, and take responsibility for the testing session.

(2) It is necessary to have a large enough number of items to accommodate the various branching patterns which might be psychometrically desirable for particular purposes. Alteration in the branching pattern should be achievable by a change in computer program, and test items should be housed in a random access device affording equal inter-item time durations.

(3) Items would be of the multiple-choice type. Their presentation would be either examinee-paced or machine-paced, with tests being either the pure speed type or the power type with administrative time limits.

(4) Simplicity of responding is essential, as are provisions which will permit examinees to omit items and to change answers, and to exercise the machine's features during an initial familiarization period.

(5) The output of a test administration should be a durable record of test and examinee identification, items attempted, vacillation among response alternatives, response latency, correctness of alternatives chosen, total raw score over all items, converted score where applicable, and certain administrative information which may be needed.

System Description

The major portions of the system consist of the examinee station, the proctor station, and the central computer.

Examinee Station. The examinee station is an enclosed private area about four feet wide by six feet long, containing a projection screen, a pushbutton panel beside the screen, and a cathode ray tube directly

---

[3] National Bureau of Standards. Report on a design study of a programed testing machine. Washington, D. C., March 1964.

below. Multiple-choice test items are displayed on the projection screen, the pushbutton panel is utilized by the examinee to select and record his answers, and the CRT presents instructions and other information to the examinee.

The screen on which the items are presented is a 15-inch square and of the rear-projection type. The projector is a Teleprompter Model RA-100 which is a high quality, random access (carousel), 35mm projector, with capacity of 100 slides.[4]

The examinee responds to the test items through use of the pushbutton panel. This is a 19-inch by 5-inch panel containing a vertical array of 9 buttons, and a 10th button offset to the left. The vertical array is directly beside the item response alternatives projected on the screen, such that each button in the array is lined up with one of the multiple-choice alternatives. Each of the buttons is also labeled to correspond to the identification letter of its adjacent item alternative. In using the pushbuttons to respond, the examinee may change his selection at any time merely by depressing another of the buttons. The single, offset, button is utilized by the examinee to record his final selection. After this button is depressed no changes may be made. The offset button is labeled RECORD.

A 10-inch rectangular cathode ray tube (6-inch by 8-inch viewing area) is located directly below the panel/projection-screen assembly. This CRT presents feedback of the identification letter of each response alternative selected, administrative instructions, remaining time, and a total score. Items have programed time limits, for psychometric or administrative purposes, and the examinee's remaining time on each item is presented digitally in 5-second intervals. Figure 1 shows the examinee displays and pushbutton controls in detail.

For troubleshooting purposes an auxiliary item of equipment, the Digital Control Unit (DCU), has been included in the system and is located in a remote area of the examinee station. The DCU presents to a technician or proctor the number of the slide that the computer has called up, for comparison with the current display.

Proctor Station. The proctor station consists of one console, which contains a typewriter keyboard and CRT identical to that at the examinee station. The proctor performs three functions: he assigns examinee identification, he monitors the examinee, and he restarts a test in the event of an unusual stoppage. The assignment of examinee identification

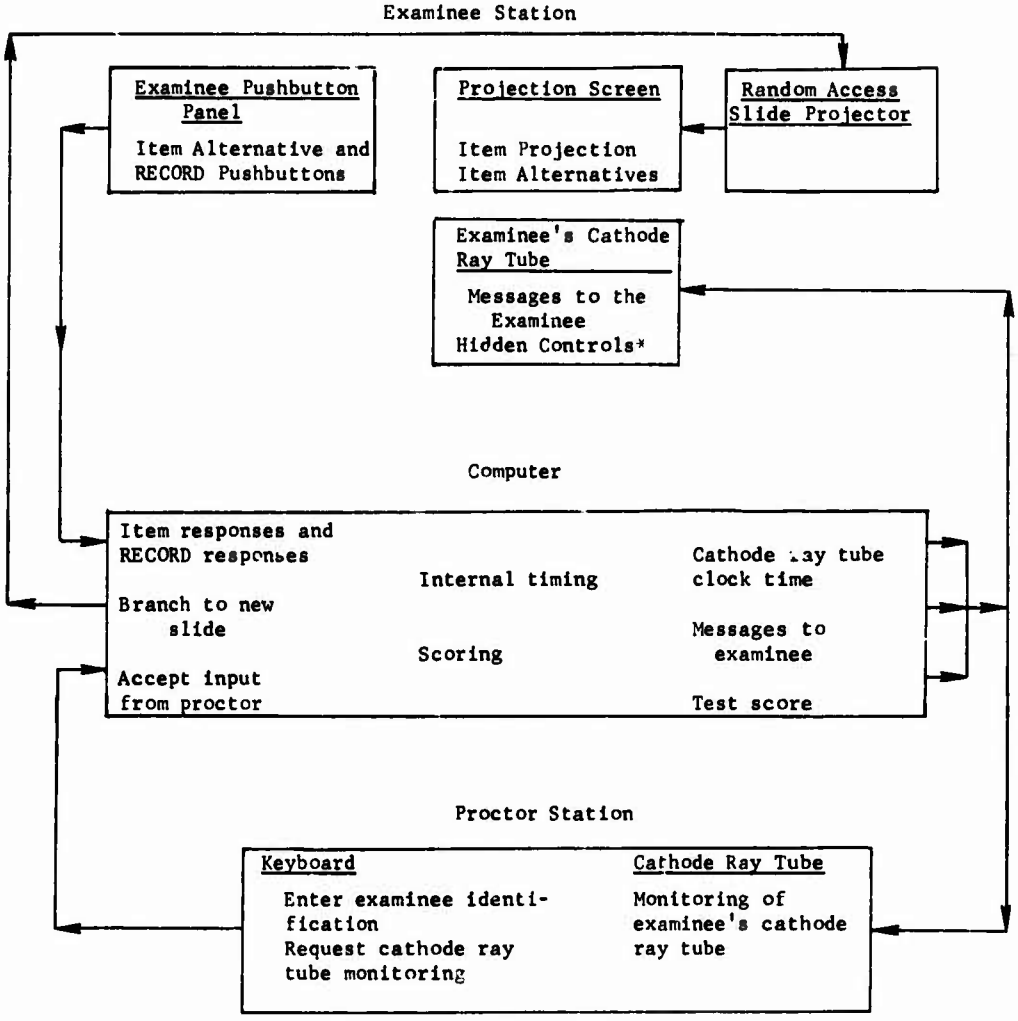
[4] Commercial designations are used only for precision of description. Their use does not constitute endorsement by the Army or ARI.

Figure 1. Examinee station

utilizes the typewriter keyboard, the monitoring utilizes the CRT, and the unusual-event restarting functions require the proctor to utilize hidden controls at the examinee station. Although only one examinee station has been built to date, the capability is present for a proctor to elect to monitor the display on any of several examinee CRTs merely by typing a display identification code.

Computer. The ARI computer, which controls and directs all test administration, is a CDC 3300. This is a high-speed, general purpose digital computer, with 65,000 words of core storage. It is a larger and faster machine than the testing system requires, but was utilized because of its availability. Programed testing takes advantage of the time-sharing capability of the central processor in the real-time mode, and retains outputs on disk pack for subsequent printing and/or punching of cards. These outputs include examinee identification, item identification, all examinee responses, latencies, and total score(s). More than one examinee and proctor station can be controlled simultaneously by this computer.

Figure 2 is a diagram of the entire system, indicating relationships among the major components and functions of the equipment.



**Hidden controls at the examinee's cathode ray tube provide the proctor the capability to restart the test in the event of an unusual stoppage.*

Figure 2. Schematic representation of the programed testing system

COMPUTER PROGRAMING

Two kinds of programing elements are used in the system: (1) subroutines used to communicate with the laboratory devices (for example, a subroutine which functionally blanks a CRT screen), and (2) the program and subroutines relevant to the test administration itself. The former are pre-existing programs in the Army Research Institute library. The programing specific to the testing system has the functions of selecting items and instructions to be displayed, displaying them on the projection screen and CRTs, accepting examinee's responses, scoring items, and providing for data storage. The steps in executing this sequence are summarized in Figure 3 and described in detail below.

The first programing step is to instruct the proctor to identify the examinee to the computer. The proctor's CRT is illuminated with a message requesting this information. When the proctor has entered the information via his keyboard, he leaves his station and ushers the examinee to the testing station. The program has displayed general instructions on the examinee CRT at the test station, and the examinee may ask the proctor any questions at this time. When the examinee understands what he is to do, the proctor leaves. From this point on, the testing system is fully automatic with the computer reacting to all examinee responses.

The examinee indicates to the computer that he is ready to start the first test item by depressing a pushbutton on his panel. A few practice items will generally be administered, followed by actual test items. The examinee is permitted to work at his own pace, but the program does not allow an excessive amount of time for responding to an item. An administrative time limit is part of the program and, as each item is presented on the projection screen, the number of seconds remaining to that limit is presented on the CRT.

The computer is signaled of examinee responses through the examinee's depression of pushbuttons at his panel. One of the nine lettered buttons will normally be depressed, followed by the RECORD button. The program directs computer storage of the letter alternative chosen, and feedback to the examinee on his CRT. The response is scored by the computer, and the result is utilized for selection of the next item to present. Under most conditions of branching tests, if the examinee answered correctly the next item would be more difficult, if he answered incorrectly the next item would be less difficult. Separate subroutines have been prepared so that the change to a different branching strategy (or item pool or instruction set) can be accomplished with only a change to a subroutine while leaving the master program intact.

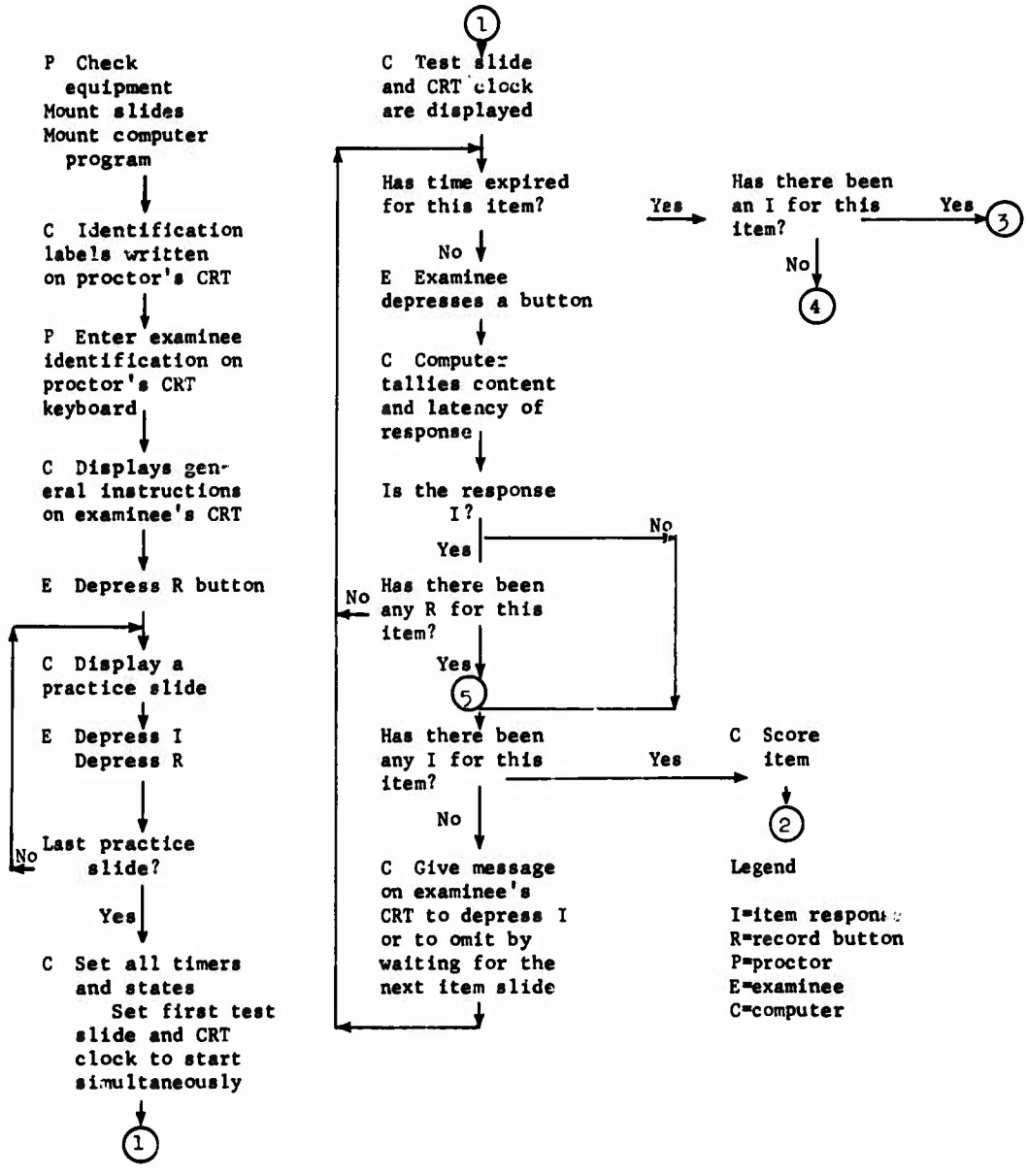


Figure 3. Flow chart -- operation of a single testing station

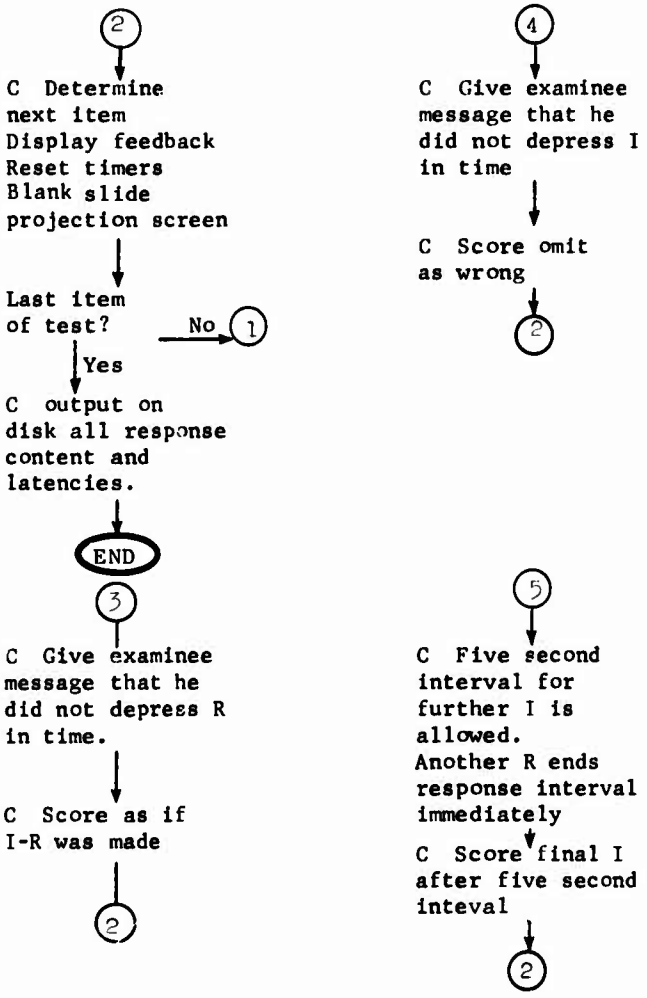


Figure 3. Flow chart -- operation of a single testing station (continued)

Response Analysis

The most common response sequence is the one described above, that is, the examinee's selection of an item alternative (I) followed by the RECORD button (R). This sequence is symbolized as I-R.

Nine other response patterns are provided for in the program. The first of these is vacillation among item alternatives, symbolized I_1 . . . I_7- R_1. This is the circumstance of the examinee changing his mind and selecting a different alternative. An examinee may depress as many as seven item alternative buttons before depressing R. If he should depress an eighth different button the assumption is made that he either doesn't understand the instructions or is not taking the test seriously, and the testing is stopped. The proctor may restart the test by means of a normally hidden button at the examinee station, which restarts the testing and retains all previous responses.

A second unusual response pattern is the repeated depression of the same item alternative pushbutton. This is treated in the same way as the pattern above, as are depressions of any combinations of seven of the same or different pushbuttons before depressing RECORD.

A third unusual response pattern is the depression of I followed by a long wait before depressing R. Fifteen seconds after depression of I a CRT message reminds the examinee that he has not depressed R. In most cases R will merely have been overlooked by the examinee, and he will depress it after the reminder (I-pause-R).

Fourth, if the examinee fails to depress R after I within the total administrative time limit for the item (that is, I-only), the response is scored as though R had been depressed, and a message so informs the examinee.

Fifth, if the examinee reverses his button depressing, i.e., depresses R before I, that I becomes the response of record, just as if it had been depressed in correct sequence. In this reverse sequence, the examinee is still permitted to change his mind and depress another item alternative pushbutton. He is allowed five seconds after the first I, and a subsequent I within this span becomes the one of record. This sequence (R-I_1, . . . I_7), is just a special case of the reversed sequence (R-I).

Two responses are scored as omissions. One is the case of no pushbutton being depressed, the other is the case of R-only. The former occurs when the administrative time limit for the item has been exceeded. The examinee is informed via CRT screen message that he did not answer in time. To minimize the inadvertent occurrence of this type of response, a "fifteen seconds left" message is flashed on the CRT screen if there has been no response to that time. In the case of the R-only response, a special message is presented on the CRT to alert the examinee to make a selection, that his response is incomplete. If the examinee still does not depress an I, the response is scored as an omission.

Responses made during the interim between slide (item) projections do not get entered into the system, and elicit no feedback to the examinee.

Finally, the system will not accept multiple responses to an item. If two item pushbuttons are depressed simultaneously (which should occur very rarely because of the very short resolving time of the electronic equipment), one of the two will be selected or a totally foreign symbol will be generated. In the first case the feedback to the examinee's CRT will indicate the choice that was made, which he may accept or override; in the second case the CRT screen will be blanked and the examinee can enter a selection.

SUMMARY OF PROGRAMED TESTING SYSTEM CAPABILITIES

Five broad requirements for the programed testing system were identified in the second section of this report. This section correlates the specific capabilities of the system with those requirements.

Requirement 1: Fully automatic system.

Except for the proctor's duties the system is fully automated. In an operational setting a single proctor could command several stations simultaneously. Discussion has covered all of the possible response styles which might be encountered and the system's automatic handling of these in continuing the administration of the test. Timing factors are readily modified so that neither unnecessary speeding of the test occurs, nor are long waits caused by a slow examinee.

Requirement 2: Random item access, large number of items, and presentation in any programed order.

The system utilizes a carousel tray which can accommodate 100 slides. For the most customary type of branching schedules this affords 10-15 item tests or subtests. Multiple carousels, each containing one test of a battery, are conceivable at an administration-time cost of no more than a short break for the examinee.

Requirement 3: Multiple-choice items, several response alternatives per item, possible self-pacing.

The system accommodates as many as nine response alternatives per item. Complete self-pacing takes place within the framework of administrative time limits.

Requirement 4: Simplicity, and provision for change of response and for omission.

Omission and change of response are accommodated. Responding is obvious and simple; in fact, responding is more straightforward than it is with conventional tests and answer sheets. Since response pushbuttons are physically aligned directly beside lettered item alternatives there is negligible risk of examinee clerical errors in matching item alternative to a letter label; and the examinee is also provided immediate feedback of the letter label via his CRT.

Requirement 5: Complete output record, in durable form, of all identifying information, responses, latencies, and total scores.

A single page of computer printout provides all of the required information for each examinee, and the system memory can accommodate as many as 40 examinees per day at each testing station administering short branching tests.